



A methodological investigation of the Intermodal Preferential Looking paradigm: Methods of analyses, picture selection and data rejection criteria



Claire Delle Luche^{a,*}, Samantha Durrant^b, Silvana Poltrock^{c,d}, Caroline Floccia^e

^a Department of Languages and Linguistics, University of Essex, United Kingdom

^b University of Liverpool, United Kingdom

^c Université Paris Descartes, Sorbonne Paris Cité, Paris, France

^d CNRS, Laboratoire Psychologie de la Perception, UMR 8158, Paris, France

^e Plymouth University, United Kingdom

ARTICLE INFO

Article history:

Received 18 August 2014

Received in revised form 26 May 2015

Accepted 30 May 2015

Available online 10 July 2015

Keywords:

Intermodal Preferential Looking

Time course analysis

Methodology

Word recognition

ABSTRACT

The Intermodal Preferential Looking paradigm provides a sensitive measure of a child's online word comprehension. To complement existing recommendations (Fernald, Zangl, Portillo, & Marchman, 2008), the present study evaluates the impact of experimental noise generated by two aspects of the visual stimuli on the robustness of familiar word recognition with and without mispronunciations: the presence of a central fixation point and the level of visual noise in the pictures (as measured by luminance saliency). Twenty-month-old infants were presented with a classic word recognition IPL procedure in 3 conditions: without a fixation stimulus (No Fixation – noisiest condition), with a fixation stimulus before trial onset (Fixation, intermediate), and with a fixation stimulus, a neutral background and equally salient images (Fixation Plus – least noisy). Data were systematically analyzed considering a range of data selection criteria and dependent variables (proportion of looking time towards the target, longest look, and time-course analysis). Critically, the expected pronunciation and naming interaction was only found in the Fixation Plus condition. We discuss the impact of data selection criteria and the dependent variable choice on the modulation of these effects across the different conditions.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last four decades, a considerable amount of energy and creativity has been devoted to designing and testing numerous experimental methods to investigate early speech perception and language comprehension in young children. One of the most popular methods is the head-turn preference paradigm (Polka & Bohn, 1996; Werker, Polka, & Pegg, 1997; Werker & Tees, 1983, 1984) which is primarily used with infants aged from 5 to 16 months of age to investigate listening preferences and discrimination. The study of word recognition or word learning from the age of 12 months (Schafer & Plunkett, 1998) relies on two paradigms, the Switch task (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002) and the Intermodal Preferential Looking paradigm (IPL, Bailey & Plunkett, 2002; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Swingle & Aslin, 2000), also called looking-while-listening procedure (Fernald, Zangl, Portillo, & Marchman, 2008).

* Corresponding author.

E-mail address: chdell@essex.ac.uk (C. Delle Luche).

In principle the IPL procedure is more versatile than the Switch task, whereby a new label is presented alongside a new visual item until a looking time threshold is reached, followed by a trial where the label is maintained or replaced by another (this is the switch trial). Indeed, whereas the Switch task is designed for novel word learning situations, the IPL allows for the investigation of novel word learning (e.g., Gurteen, Horne, & Erjavec, 2011; Schafer & Plunkett, 1998; Swingley & Aslin, 2002, 2007) together with familiar word recognition (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Fernald et al., 2008; Houston-Price, Mather, & Sakkalou, 2007; Mani, Coleman, & Plunkett, 2008; Mani & Plunkett, 2007, 2008, 2011a; Ramon-Casas, Swingley, Sebastián-Gallés, & Bosch, 2009; White & Morgan, 2008), mutual exclusivity (Houston-Price, Caloghiris, & Raviglione, 2010) and most recently has been adapted for priming tasks (e.g., Arias-Trejo & Plunkett, 2009, 2013; Mani, Durrant, & Floccia, 2012; Styles & Plunkett, 2009).

The standard procedure consists in presenting pairs of images horizontally on a screen for several seconds and, mid-trial, playing a target word or a carrier sentence. A trial is thus divided in a pre-naming and a post-naming phase (for longer post-naming windows, see Zangl, Klarman, Thal, Fernald, & Bates, 2005). For the duration of the experiment, eye movements are recorded by cameras mounted above each image (or eye-tracking when available). These eye movements are then time-locked onto each trial and traditionally manually coded frame by frame. With the growing use of eye trackers, gaze coding is now often automatic.

Fernald and colleagues (2008) provided a comprehensive review of the procedure and improvements added progressively since the first introduction of the paradigm, and listed factors that need to be controlled: both images have to be matched for size and visual salience; auditory stimuli have to be controlled for duration across items; target side has to be counterbalanced overall. One of their recommendations is that across all participants both objects in a given trial should be used as target and as distracter, since it is the best control to avoid any preference for one stimulus over another. Although desirable, such a control is not always possible given the restricted choice of stimulus items in young children and the need for a sufficient number of trials per participant. Some experiments have controlled for this possible preference effect (Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002), presenting the same visual stimuli at least twice, while others have not (Durrant, Delle Luche, Cattani, & Floccia, 2014; Floccia, Delle Luche, Durrant, Butler, & Goslin, 2012; Mani et al., 2008) but still found comparable results. One possible compromise, as suggested by Fernald et al. (2008), is to ensure an equal preference for both pictures during the pre-naming phase by monitoring looking times in silence during a pilot experiment. However, to our knowledge, such a pretest or control for the absence of a pre-naming image bias is not reliably reported in the literature (with the exception of the results section in Swingley, Pinto, & Fernald, 1999). Another way of controlling for pre-naming visual preferences is to take into account looking behaviour in the pre-naming phase in the statistical analyses, which is frequently reported (e.g., Mani & Plunkett, 2007; Meints, Plunkett, & Harris, 1999).

Finally, Fernald et al. argue that, contrary to adult visual experiments, a central fixation point right before naming is not necessary as children would not follow such an implicit instruction (note that White & Morgan, 2008, used a centering stimulus before the pre- and the post-naming phases, while Gurteen et al., 2011, used a centering light before post-naming).

Despite the excellent review by Fernald et al. (2008) the relatively recent addition of the IPL paradigm to the field of developmental psycholinguistics means that researchers often face choices regarding the procedure itself and the methods of analyses, all of which can have important consequences on the observation of an experimental effect. For example, as demonstrated by Arias-Trejo and Plunkett (2010), choosing a distracter image which is perceptually close to the target image (e.g., a balloon paired with an egg) can result in unwanted interference effects so that 18- to 24-month-olds fail to identify the target image. The consequences of other methodological choices (such as the use of a fixation point) on the robustness of the experimental effects are largely unknown. As we will show below, a review of the recent literature reveals a great deal of variation in many aspects of the procedure, as well as in the selection of the dependent variables used for the analysis of looking times. Appendix A provides details on methodological aspects such as the presence of a central fixation stimulus or the duration of the pre- and post-naming phases across a range of studies that have used the IPL methodology. The aim of the current study is to complement and extend Fernald et al.'s review by examining how the different methodological choices and the different methods of looking time analyses impact on the observation of significant results. In three experiments testing familiar word recognition with 20-month-olds, we manipulated the level of visual noise (with saliency as measured by luminance and the presence of a central fixation point), and provided a systematic and thorough analysis of looking times using those methods most representative of the current literature. The main objective of this paper is to provide researchers with some data-grounded recommendations about the best practices when using the IPL procedure.

1.1. Central fixation point

A review of the literature using the preferential looking paradigm shows that around half the experiments use a centering stimulus at the beginning of each trial, visual or auditory (Curtin, 2010; Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008; Meints et al., 1999; Meints, Plunkett, Harris, & Dimmock, 2002; Meints, Plunkett, Harris, & Dimmock, 2004; Schafer & Plunkett, 1998, see Appendix A) while the other half do not report such a practice (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Mani et al., 2008; Mani & Plunkett, 2007, 2008). This stands in contrast with ERP studies, and adult experiments generally, where a fixation stimulus is systematically presented to centre the participant's attention before trial onset (Kuipers & Thierry, 2011). When a central fixation point is used, trials are always triggered by the experimenter, but not systematically when such cue is not used, in which case trials are sometimes automatically interspaced (Ramon-Casas et al., 2009; Swingley, 2003, 2007).

Although it is not possible, from these studies, to draw direct comparisons between results obtained with and without fixation points given the variety of investigated topics, one can estimate that centering attention, even furtively, should benefit the procedure and the quality of the data, especially since it ensures that the child is attentive to the screen immediately before trial onset. The presentation of an image in the centre of the screen after termination of a trial has multiple advantages: (i) since the trial is triggered only if the child is looking at the centre, it ensures the child is attentive and active; (ii) attention is attracted back to the middle of the screen, giving the same weight to the probability that the first look will be at the target or the distracter once the trial begins; (iii) looking at the centre right before trial onset should encourage the child to explore all new stimuli that appeared in her peripheral vision. This is of particular importance when considering that trials where the child does not look at both images in the pre-naming phase can be discarded in the statistical analyses (e.g., [Mani & Plunkett, 2007](#)); and (iv) by having something to look at for the whole duration of the experiment, the entire procedure becomes dynamic and eventful, maintaining the child's interest. One of the aims of this study will be to verify if looking behaviour is affected by the potential noise reduction provided by a fixation point in a classic IPL task.

1.2. *Quality of visual stimuli*

[Fernald et al. \(2008\)](#) recommended controlling the visual stimuli for size, animacy and salience (in the sense of visually engaging images, especially by matching objects for animacy). Regarding saliency, experimenters decide on the pictures without any objective measures. Visual stimuli are often static colour photographs of objects on a white or grey background (e.g., [Mani & Plunkett, 2007](#); [Swingley & Aslin, 2000](#)), sometimes a mix of realistic drawings and photographs ([Swingley, Pinto, & Fernald, 1999](#)), or quite exceptionally line drawings, coloured ([White & Morgan, 2008](#); [White, Morgan, & Wier, 2005](#)) or plain (with made up animals, [Mather & Plunkett, 2011](#)). So as to enhance interest in the visual stimuli, pictures sometimes move in synchrony on a vertical axis ([Swingley, 2003](#); [Swingley & Aslin, 2002, 2007](#)). In word learning studies, made up objects (obtained by editing colour photographs as in [Schafer & Plunkett, 1998](#)) are visually comparable to real objects. To our knowledge, only [Gurteen et al. \(2011\)](#) have presented real objects to the participating infants.

Such variability in the selection of visual stimuli in the literature has been enabled by the possibility of retrieving images and photographs from the internet, departing from the perceptually controlled line drawings from [Snodgrass and Vanderwart \(1980\)](#) or their coloured version ([Rossion & Pourtois, 2004](#)) to achieve more naturalistic representations. One way to control for the quantity of information provided by photographs is to remove any background, even though the percept is less naturalistic looking. This practice is supported by [Meints et al. \(2004\)](#) who showed that background affected word recognition. Younger children (15 months) do not recognize a sheep when it is pictured with a naturalistic background (e.g., a sheep on grass), while they do when the sheep is presented without background or on an unusual – or less rich – background. Older children recognize the sheep regardless of background, although the distracting effect of the typical background was still observed to a certain extent.

Perhaps estimating the basic visual salience of the stimuli would be another, quantifiable, step towards ensuring that target images are not more attractive than their corresponding distracters, to complement experimenter judgments. Note that this is a purely visual control of salience, away from the subjective salience discussed by [Fernald et al. \(2008\)](#) or the more cognitive saliency maps (for a review, see [Althaus & Mareschal, 2012](#)). This will be achieved in the current study by performing cross-correlations of the pairs of images presented ([Chinga & Syverud, 2007](#)). Images are transformed into a matrix containing the luminance of each pixel, then into a vector. The cross-correlation compares then the two vectorized images: a high correlation score will mean that the two images are similar in salience. By contrasting a more or less visually noisy set of pictures, we will examine its effect on looking behaviour.

1.3. *Methods of analysis*

Regardless of the task the participant is engaged in, all experiments in the IPL literature divide an experimental trial into a pre- and a post-naming phase, based usually on the onset of the target word (with the exception of priming studies in which there is no pre-naming phase, e.g., [Arias-Trejo and Plunkett, 2009](#)). The selection of analysable data as well as the choice of dependent variable varies according to research groups and, on occasion, differs for a single experiment (see [Appendix A](#)). The most typical time window of interest starts 367 ms after word onset (and less frequently word offset): indeed it has been established that a minimum of 233 ms is necessary to obtain stimulus-related saccades, and that this latency depends on age, vocabulary size or task complexity ([Fernald et al., 2008](#); [Fernald, Perfors, & Marchman, 2006](#); [Zangl & Fernald, 2007](#); [Zangl et al., 2005](#)). This time window usually ends at 2000 ms, as it is generally considered that later looking behaviour is no longer related to the processing of the auditory stimulus. When plots of the time course for proportions of looks to the target are included, usually at end of the result section ([Arias-Trejo & Plunkett, 2010](#); [Fernald et al., 2008](#); [Swingley & Aslin, 2000](#)), the time window can then be justified a posteriori, the visual inspection confirming that roughly 2000 ms after word onset, looking behaviour resumes to chance level (that is, equal looks to the target and the distracter). However, since latency is a function of at least age ([Zangl & Fernald, 2007](#)) and vocabulary size ([Fernald et al., 2006](#)), the whole looking behaviour can also be influenced by task difficulty. This is the case for example when mispronunciations of words are minor ([Mani et al., 2008](#); [Mani & Plunkett, 2011a](#); [Swingley, 2007](#); [Swingley & Aslin, 2002](#)). A systematic time window, regardless of data distribution in the post-naming phase, may overlook meaningful looks if children are still looking at the target after 2000 ms. It seems recommendable (see [Fernald et al., 2008](#)) that the first, and not the last, step in data analysis should be

the systematic plot of the unfolding looking behaviour, so as to ascertain that the window of analysis comprises all the word processing and task related looks. This is common practice in EEG or MEG experiments, since electrophysiological markers can vary in location and time period (e.g., Bastiaansen, van der Linden, ter Keurs, Dijkstra, & Hagoort, 2005; Hagoort & Brown, 2000). Perhaps a way to enhance the precision of the results would be to determine statistically the exact time window when the two conditions differ (target vs. distracter in simple naming tasks, correct vs. incorrect pronunciation in mispronunciation experiments). To our knowledge, with the IPL paradigm, only one instance of such an analysis has been published so far (von Holzen & Mani, 2012, see Maris & Oostenveld, 2007 for a detailed explanation); the authors showed that in addition to standard comparisons of looking times averaged across pre- and post-naming trials, it was possible to identify accurately a specific time window where performance between conditions differed (in their case, 1140–1580 ms after target word onset). There are two advantages for this method: first, representing time course plots gives a dynamic evaluation of visual/linguistic processing; secondly, this data driven method is objective and prevents a priori judgement of the data.

In relation to the looking behaviour, two types of dependent variables are usually considered: proportion of target looking (taking into account, or not, the pre-naming phase, see Appendix A), and latency of shifts to the target. They are respectively assimilated to a correct response and a reaction time (see Fernald et al., 2008). Other measures have also been used, and usually provide comparable direction of results, such as total looking time (that is, the sum of looks to the target during the post-naming phase minus those to the distracter), or the longest look (longest single fixation to the target).

Data filtering or pre-processing is where the greatest variation across experiments is observed (see Appendix A). In word recognition tasks, with or without a mispronunciation element, words are selected so that they are likely to be known a priori by all participants according to standardized norms (thus reducing considerably the number of potential stimuli, re. Ramon-Casas et al., 2009; Swingley et al., 1999), or by at least 50% of children of the corresponding age (e.g., Styles & Plunkett, 2008). Then, some authors further filter the data by analyzing only trials where parents report the words as known by the child (e.g., Mani & Plunkett, 2011b). Whether it is necessary or not to check for infants' knowledge of distracter will be addressed here. On the one hand, if the child does not know the distracter, she might be looking more at the target once it has been named simply because the target object is the only object for which she has a name, and not because she recognizes the link between the label and this object; this would artificially inflate the target looking time. On the other hand, a child who does not know the distracter's name would look longer at its picture in mispronunciation trials, in the spirit of the study by White and Morgan (2008) whereby unknown objects were presented as distracters. This may be an advantage for strengthening mispronunciation effects. The impact of filtering out trials where the child does not know both the target word and the distracter will be evaluated in the current study.

The criteria used to select the attended trials is also variable, with some considering only long enough fixations to the images (1500 ms in each phase, Bailey & Plunkett, 2002), or fixation to both images in the pre-naming phrase or at least, throughout the trial (Mani & Plunkett, 2007).

Finally, data cleaning is achieved by excluding participants not contributing to all experimental conditions (Fernald et al., 2006; second analysis in Styles & Plunkett, 2008), or whose data points fall outside normality (Fernald et al., 2006; see also Mani & Plunkett, 2007). Although all these types of pre-processing or filtering allow for cleaner datasets, comparability across experiments would benefit from consistent practice, preferably on the measures that are the most conservative. Falling on an agreement on exclusion criteria or on the best-suited age-specific time window would be desirable. Indeed, some analyses only include trials where the child is looking at a picture (final analysis in Fernald et al., 1998), while others reject children not contributing to all conditions (e.g., Fernald et al., 2006; second analysis in Styles & Plunkett, 2008), set a minimum looking time (e.g., Bailey & Plunkett, 2002; Ballem & Plunkett, 2005) or only include trials where both pictures are fixated (Mani & Plunkett, 2007, 2008).

The goal of the present research is to examine how the different methods of data analyses are resistant to methodological alterations, or noise, such as image quality and the presence of a central fixation stimulus. For this purpose, we ran three versions of a classic IPL procedure testing the detection of mispronunciation of familiar words, varying the pictures' saliency and background, and the presence of a fixation point. For each experiment, we evaluated how the degree of visual noise, together with the different criteria for data selection, modified the different dependent variables.

Here, the stimuli were comparable to those in Mani and Plunkett (2007), in which 15-to-24-month-olds were presented with two images on both sides of a screen and heard, mid-trial, correctly pronounced in a carrier sentence such as "Look, dog!" for half of the trials, or as a mispronounced version of the target word ("Look, bog!") in the other half of trials. If children recognize lexical entries of familiar words only if they are pronounced correctly (as is expected from the age of 18 months, e.g., Mani & Plunkett, 2007), a naming effect should be observed in correctly pronounced trials, but not (or significantly less) in mispronunciation trials resulting in an interaction between naming and pronunciation – the key result in these studies. The participants in the current study are 20 month olds, and so results comparable to 18 month olds can be expected, if not stronger, because their lexical repertoire increases steadily and their phonological sensitivity seems stable around these ages (for a developmental Switch task, see Werker et al., 2002).

As presented earlier, we hypothesize that adding a fixation stimulus between trials should enhance the quality of the data. Indeed, since experimental trials are then only triggered when the child is attentive to the screen, post-hoc measures of attentiveness (by checking the videos, re. Fernald et al., 2008; or excluding trials with looks <1500 ms, re. Bailey & Plunkett, 2002) are less critical yet desirable. The added advantage of having a centred fixation stimulus is that it should entice the participants into looking at the two images that appear in their peripheral vision at trial onset.

The second manipulated methodological choice is the uniformity of picture background and the choice of visual stimuli. As tested by Meints et al. (2004), a typical background is distracting and reduces the naming effect in younger children, and to a certain extent in 18-month-olds. To capitalize on these findings and fully appreciate the extent of any distraction relating to a naturalistic background on the naming effect, we compared two sets of images, one where the content was naturalistic (mostly with a background), and one where there was no background, leaving the stimuli in isolation. In addition, we examined the effect of image saliency on looking behaviour, especially in the pre-naming phase, which, to our knowledge, has never been investigated experimentally in infants, despite recommendations to control for it (Fernald et al., 2008).

To sum up, we manipulated the presence/absence of a central fixation point together with the uniformity of picture background and picture saliency, to evaluate the effect of experimental noise in an auditory word recognition task. In the first condition (No Fixation), no central fixation stimulus was used and no attempt was made to control for the picture background colour or saliency. This is the noisiest condition. In the second condition (Fixation, intermediate noise condition), a central fixation point was added and the same images were used as in the No Fixation condition. In the third condition (Fixation Plus), the central fixation point was augmented by a systematic absence of background colour for pictures and target-distracter pairs were selected so that their saliency was highly correlated. This is the least noisy condition.

We expect that adding the fixation stimulus should encourage more looks to both images in the pre-naming phase, leading to cleaner data in terms of trial rejection, and more balanced looks towards target and distracter (as seen for example by shorter “longest look” measures). The adjunction of better controlled image background and saliency should also contribute to enhance the identification of objects, reducing the “back and forth” between target and distracter, especially during the post-naming phase. This would translate into longer “longest look” measures but also in a larger naming effect in correct trials.

To fully estimate the impact of experimental stimuli manipulations on data quality, and therefore get a clear picture of the robustness of the method, we will provide different types of analyses. First, we will evaluate results for the classic 367–2000 ms post-naming time window and will examine effects of conditions on the different dependent variables used in the literature. We will also look at the impact of the different criteria used for trial rejection on the results. Then we will present a newer type of analysis that takes the time course of looking behaviour into consideration and thus avoids averaging over the whole post-naming window (von Holzen & Mani, 2012). While time course is seemingly the crucial aspect of the IPL paradigm or other visual paradigms, it has been rarely exploited in IPL experiments yet should provide with informative and complementary results.

2. Methods

In this experiment the classic mispronunciation IPL paradigm is used: two objects are presented side by side on a screen and one is named halfway through the trial. Pronunciation is correct for half the trials (e.g., “Look! Bed!”), and incorrect for the other half (“Look! Bud!”), for *bed*). We developed three versions. In the No Fixation condition, no central fixation stimulus was used and images were simply controlled for suitability and size but not for background, colour or perceptual saliency (they mostly had a naturalistic background). In the second, Fixation condition, a fixation stimulus was added before the start of every trial, and in the last condition, Fixation Plus, this was augmented by controlling the background colour of the pictures and their relative saliency.

2.1. Participants

Participants from the three conditions were matched for gender, age and vocabulary scores as measured by the Oxford Communicative Development Inventory (Hamilton, Plunkett, & Schafer, 2000).

Twenty 20-month-olds were successfully tested for the Fixation condition (mean age 20;07; range 19;19 to 21;13; SD 3 days; 8 females and 12 males). Average OCDI scores were 241 words (SD = 19.1) in comprehension. One additional participant was excluded for fussiness.

The population in the Fixation condition served as the baseline for the selection of participants in the two other conditions as they were matched on their OCDI scores in comprehension.

Thus, 20 children out of an initial group of 39 children constituted the No Fixation condition ($M = 19;16$; range 18;20 to 20;29; SD 4 days; 8 females and 12 males). Their comprehension score was 234 words on average (SD = 17.7). Participants for the Fixation Plus condition ($M = 19;27$; range 18;14 to 21;5; SD = 5; 8 females and 12 males) were selected from the participants of Durrant et al. (2014, 32 toddlers), with an average comprehension score of 241 words (SD = 15.6).

2.2. Stimuli

The stimuli were 32 monosyllabic consonant initial words taken from the OCDI (see Appendix B), half targets and half distracters. All are imageable nouns. Target words were judged as known by at least 40% of 20-month-olds (database from the Oxford Babylab), and they were paired with a distracter sharing the same onset consonant and phonemic structure (e.g., dog and duck). Out of the 16 targets, participants heard 8 labels that were correctly pronounced and 8 that were mispronounced. Following Mani and Plunkett (2007), mispronunciations were obtained by changing one phoneme on one

or more dimension, either on the onset consonant or the vowel (4 trials each), leading to a pseudoword or a very rare word in infant-directed speech (e.g., *bud*).

The speech stimuli were recorded in an enthusiastic child friendly manner by a female native speaker of British English. Recordings were conducted in a sound-attenuated booth, digitized at a rate of 44.1 kHz and a resolution of 16 bits. The recorded tokens were matched so that the duration, amplitude and F0 of the correctly pronounced labels and their mispronunciations did not differ significantly.

The tokens were then spliced onto a carrier sentence “Look! *Target word!*”, with onset of the test word starting 2500 ms into the trial. Note that the auditory stimuli are identical across conditions.

The visual stimuli were photographs of the targets and distracters retrieved from the web and judged by the experimenters as good exemplars of the chosen categories. For the Fixation and Fixation Plus conditions, a smiley face was presented between trials at the centre of the screen until fixated by the participant, and followed by the next trial (triggered by the experimenter). For the Fixation Plus condition, image quality was manipulated: as exemplified in [Appendix C](#), we systematically removed background colour and matched pictures (pairs of target and distracter) for visual saliency. This was checked with Pearson’s correlation conducted on the vectorized image pairs ([Chinga & Syverud, 2007](#)). The saliency of target-distracter pairs presented to the No fixation and Fixation conditions was not well correlated (0.13 on average, for the absolute value of the *R*s), whereas a reasonable correlation for the Fixation Plus condition was found (0.35 on average, which is a good correlation magnitude, [Hemphill, 2003](#)).

Images were projected onto a screen 1.20 m away from the child, each of the stimuli image measured 52 cm diagonally and were separated by 43 cm, so that both images were comprised within 48° of visual angle with 10° of gap between them. The smiley face was centred and measured 14 cm diagonally.

2.3. Procedure

After written consent was obtained, both the participant and the parent were invited into the room set up for the IPL. An image was presented on the screen in the dimly lit room so as to entice the child into sitting in a high chair. A short animated cartoon was played on the screen to keep her entertained and looking at the screen while the experimenter adjusted the cameras on her face. The parent sat behind the child and was asked not to intervene in any way so as not to influence the child. The experimenter could hear the stimuli being presented (so that she could also hear possible parental intervention) but could not see the screen.

Each of the 16 trials (plus two for training) were started manually by the experimenter when the child was looking at the screen (anywhere for the No Fixation condition) or at the centre of the screen, that is, at the smiley face (for the Fixation and Fixation Plus condition). Participants were presented with 8 correct labels and 8 mispronounced labels. Order of presentation, pronunciation type and side of the target were counterbalanced across participants.

2.4. Scoring

The digital scoring system developed by [Meints and Woodford \(2008\)](#) was used to synchronize videos with trial onsets. Eye movements (left image, right image, middle or away) frame by frame (40 ms) were scored by skilled coders trained by the first author and naïve to the items being presented to the participants. Each eye fixation was coded, and an independent skilled coder scored 10% of the pool of data randomly for each group. Agreement between coders was high with an intraclass correlation coefficient of 0.936 ([Shrout & Fleiss, 1979](#)).

Looking times on each image were then automatically extracted for the pre- and post-naming phases, providing the proportions of looks to the target and the distracter in the pre- and post-naming phases, as well as longest look measures and frame by frame eye position for the time-course analysis.

3. Results

3.1. Time course plot and windows of analysis

Following recommendations by [Fernald et al. \(2008\)](#), proportions of looks to the target as a function of time and pronunciation type were plotted in [Fig. 1](#) for each condition (No Fixation, Fixation and Fixation Plus). Visual inspection of the plots suggests that the usual 367–2000 ms window post-naming seemed adequate, although it looks like the naming effect extended after 2000 ms post-naming for the No Fixation and the Fixation conditions. The final time course analysis (as in [von Holzen & Mani, 2012](#)) will be the best test to determine when pronunciation effects occur.

It can be seen from the plots that at the end of the pre-naming phase, looks to the target were, on average, below the expected 50% (corresponding to no preference for targets or distracters) for the No fixation and the Fixation conditions, while the proportion was more balanced for the Fixation Plus condition. While [Fernald et al. \(2006\)](#) recommend an average of 50% of looks to the target in the pre-naming phase to avoid any bias, it is often normalized by subtracting pre-naming measures from post-naming measures (see, among others, [Mani & Plunkett, 2007](#)) or computing a salience score like in [Swingle and Aslin \(2007\)](#). We will return to this observation in the discussion.

3.2. Selection of trials

First of all, we only analyzed trials where the target was known to the participant. As children were matched for vocabulary knowledge across the three conditions, we expected the proportions retained for each condition to be comparable: 95.0% for the No Fixation condition (304/320 trials), 87.8% for the Fixation condition (281/320), and 93.4% for the Fixation Plus condition (299/320). However, a chi-square test run on the raw scores revealed a significant

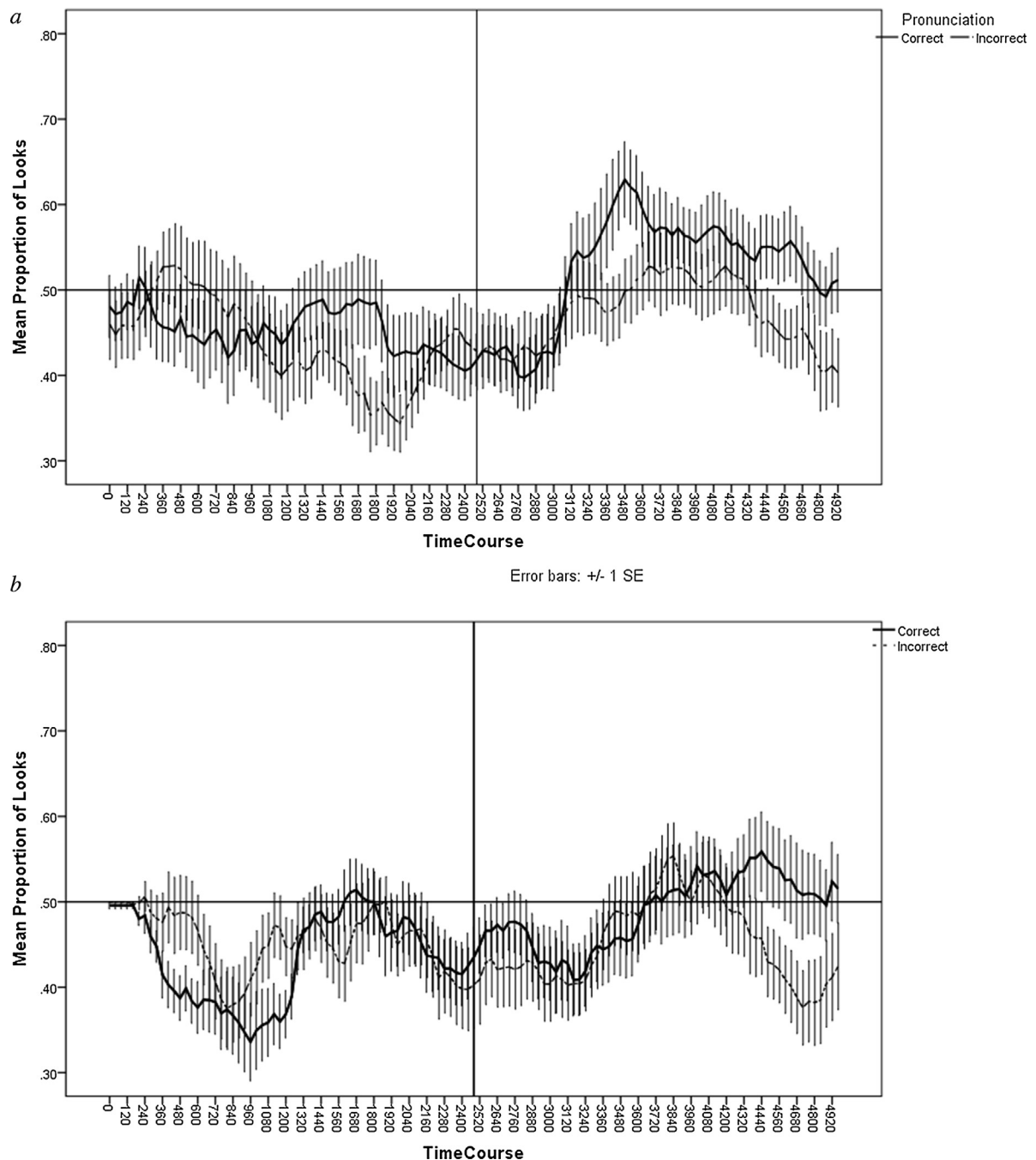


Fig. 1. Time course plot (in ms, with SD) of the PTL to the target for the No Fixation (a), the Fixation (b) and the Fixation Plus (c) conditions, with correct (solid line) and incorrect (dash line) pronunciations. The vertical line represents the onset of the target word and the start of the post-naming phase, and the grey rectangle the time period where the two pronunciation conditions differ. Significance as ascertained by the time course analysis technique (Section 3.4, point iii) is indicated by an asterisk.

difference between conditions ($\chi^2(2) = 12.55, p = .002$). A general linear model of the data with *vocabulary* (target known vs. unknown) and *condition* (No Fixation, Fixation, Fixation Plus) as factors confirms that more words were known in the No Fixation condition compared to the Fixation condition ($z = 3.001, p = .003$), but not between the No Fixation and the Fixation Plus conditions ($z < 1, n.s.$). This was very likely due to a sampling effect, since the Fixation group is the only one where two participants did not know 6/16 words, while in the other groups the maximum words that were unknown did not exceed 4/16. The χ^2 test excluding these two participants shows that then the three groups no longer differed ($\chi^2(2) < 1, n.s.$).

The second step in data selection often involves an inspection of looking time distribution trial per trial. The procedure presented by Fernald et al. (2008) includes a pre-screening of the recorded videos. Another way of retaining attended trials is to select trials where the child is looking at both pictures, either necessarily in the pre-naming phase (strict criterion), or at some point throughout the trial (lax criterion). The strict criterion is likely to result in cleaner data; however it could be that the lax criterion is preferable: training trials should be sufficient for the child to understand that pictures appear in

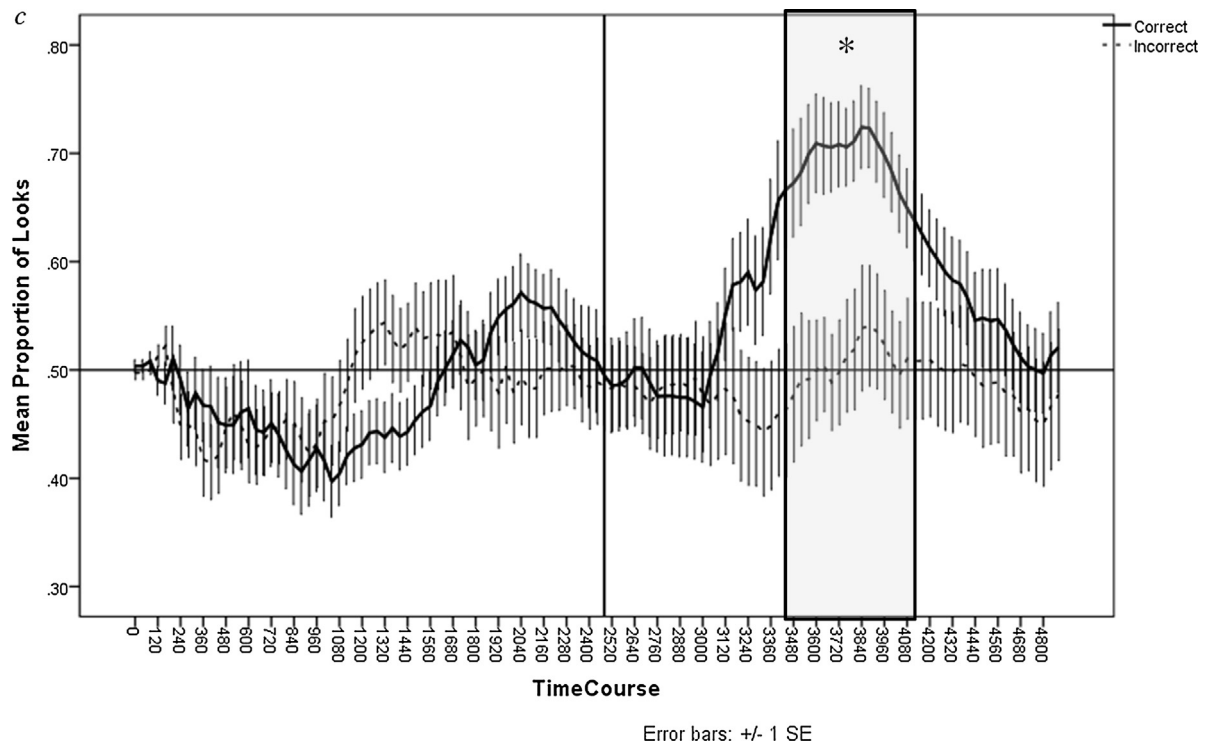


Fig. 1. (Continued).

the two corners of the screen. Consequently, upon hearing “Look! Dog!”, shifts to the target in the post-naming phase when the child did not look at the target picture in the pre-naming phase can still be considered as a sign of word recognition (re. mutual exclusivity in monolinguals, [Houston-Price et al., 2010](#)). In what follows, we present data based on the lax criterion, but we also provide in [Appendix C.1–3](#) the analyses based on the strict criterion.

With the lax criterion (looks at the target and distracter at some point throughout the trial), we retained 98.7% of the known trials in the No Fixation condition (300/304), 95.4% in the Fixation condition (268/281) and 97.0% in the Fixation Plus condition (290/299). A Pearson's chi-square test with Yates correction shows that the rejection rate did not differ across conditions ($\chi^2(2) = 4.52, p = .10$). The strict criterion (looks to both target and distracter during the pre-naming phase) retained 88.8% of the trials in the No Fixation (270/304), 90.4% of the Fixation trials (254/281) and 88.0% of the Fixation Plus trials (263/299). Again, a chi-square test revealed no main effect of condition ($\chi^2(2) < 1, n.s.$). Note that all children were included in the analysis as they contributed to all conditions (re. [Fernald et al., 2006](#)).

3.3. Looking behaviour in the pre-naming phase

Preliminary analysis of the pre-naming phase is meant to control for a potential bias towards the image that will be named later in the trial. Although averages were close to 50% during the whole pre-naming phase, they were significantly below chance for No Fixation ($M = 44.4\%$, $SD = 6.0\%$, $t(19) = -4.15$, $p = .001$) and Fixation ($M = 42.9\%$, $SD = 5.9\%$, $t(19) = -5.34$, $p < .0001$), showing a bias towards the distracter image, which was similar in both conditions ($t(38) < 1, n.s.$). In the Fixation Plus condition however this bias was no longer significant ($M = 49.1\%$, $SD = 4.5\%$, $t(19) < 1, n.s.$), and differed significantly from that in the No Fixation condition ($t(38) = -2.79$, $p = .008$) and the Fixation condition ($t(38) = -3.72$, $p = .001$).

3.4. Dependent variables

IPL studies in the literature have presented a whole range of different dependent variables (see [Table 1](#)), and included pre-naming as a factor ([Mani & Plunkett, 2007](#); [Styles & Plunkett, 2008](#)) or as a salience score ([Swingle & Aslin, 2007](#); [White & Morgan, 2008](#)). In the first case, the two resulting independent variables are *naming* (pre- vs. post-naming) and *pronunciation* (correct vs. incorrect), and in the second case only *pronunciation*. Results from the *pronunciation* effect with the salience score are then identical to the post-hoc analysis of the *naming* \times *pronunciation* interaction. As such, we will not present the salience score analysis. It is worth noticing that sometimes the pre-naming phase is not explicitly analyzed ([Fernald et al., 2008](#); [Ramon-Casas et al., 2009](#); [Schafer & Plunkett, 1998](#); [Swingle, 2003](#)).

The most widely used dependent variables are either the longest look measure (LLK) or the proportion of total looks towards the target (PTL). The PTL is usually calculated by dividing the total looking time to the target (T) by the total amount of looks to the target and distracter, that is $T/(T+D)$, in the pre- and post-naming phases. A significant increase in the PTL in the post- compared to the pre-naming phase is taken as evidence of a naming effect. A significant decrease or absence of difference will show an absence of naming effect, that is, no evidence of word recognition. The LLK, on the other hand, represents the longest single fixation on the target (and also distracter). Successful word recognition, or the naming effect, should lead to an increase of LLK in the post-naming compared to the pre-naming phase.

Descriptive statistics are presented for the analysis where children looked at both images at some point during the whole trial (lax criterion), with proportion of looks to the target (PTL, [Fig. 2](#)) and longest look (LLK, [Table 1](#)) measures for the target and distracter, in the pre- and post-naming phases.

Statistics are reported for the overall LLK and PTL as dependent variables, with *pronunciation* (correct vs. incorrect) and *naming* (pre-, post-naming) as within-participant factors, and *condition* (No Fixation, Fixation, and Fixation Plus) as between-participant factor. We also report effects and interactions of *pronunciation* and *naming* for each condition separately, as each condition could potentially constitute a stand-alone experiment.

Since the p values appear to be comparable for both measures (with the exception of the triple interaction, significant with LLK, and driven by the interaction between *naming* and *pronunciation* for the Fixation Plus condition), in the first following section we only report values for the PTL measure.

Table 1

Mean PTL to the target (in %), longest looks to the target and the distracter in the pre- and post-naming phases, for all conditions (standard error in brackets). The lax criterion was applied.

	Conditions			
	Pronunciation	No fixation	Fixation	Fixation Plus
<i>PTL to target</i>				
Pre-naming		48.3 (1.7)	42.2 (1.2)	45.5 (1.5)
Post-naming	Correct	57.5 (2.5)	47.9 (3.2)	61.9 (2.1)
	Incorrect	51.2 (3.4)	45.8 (2.9)	46.2 (3.5)
<i>LLK to target</i>				
Pre-naming		685 (27.2)	595.5 (16.9)	639.5 (23.7)
Post-naming	Correct	830.0 (46.3)	690.7 (54.8)	923.2 (42.8)
	Incorrect	752.8 (63.9)	636.4 (40.4)	640.6 (45.4)
<i>LLK to distracter</i>				
Pre-naming		736.2 (26.4)	820.6 (20.9)	750.9 (29.3)
Post-naming	Correct	591.1 (43.0)	734.1 (44.8)	528.1 (30.0)
	Incorrect	684.4 (48.5)	758.0 (51.2)	763.1 (60.7)

Detailed descriptive statistics for the other measure (LLK) can be found in Table 1, and the corresponding ANOVAs are described in the following sections. Comparisons of measures will be discussed in the next section.

3.4.1. PTL measure

For all conditions together, the statistics reveal a significant main effect of *naming* ($F(2,57)=22.63$, $p<.0001$, $\eta_p^2=.28$), with increased looks to the target after it was named (pre-naming $M=45.6\%$, $SD=1.2\%$; vs. post-naming $M=51.8\%$, $SD=1.9\%$). There was also an effect of *pronunciation* ($F(2,57)=5.40$, $p=.02$, $\eta_p^2=.09$) suggesting more looks to the target when it was pronounced correctly ($M=50.5\%$, $SD=1.6\%$) than incorrectly ($M=46.8\%$, $SD=1.6\%$). The more interesting results were the main effect of *condition* ($F(2,57)=5.73$, $p=.005$, $\eta_p^2=.17$) and the marginal interaction between *pronunciation* and *naming* ($F(2,57)=3.92$, $p=.05$, $\eta_p^2=.06$). Finally, the interaction between *condition*, *pronunciation* and *naming* was not significant ($F(2,57)=2.36$, $p=.10$, $\eta_p^2=.08$). Because the critical interaction between *pronunciation* and *naming* was marginally modulated by condition, we first repeated the analyses in each condition separately, for ease of comprehension.

In the No Fixation condition, there was a marginal effect of *pronunciation* ($F(1,19)=4.00$, $p=.06$, $\eta_p^2=.17$), and a significant effect of *naming* ($F(1,19)=6.37$, $p=.021$, $\eta_p^2=.25$), due to more looks to the target in the post-naming phase than in the pre-naming phase (pre: $M=47.6\%$, $SD=2.1\%$; post: $M=55.1\%$, $SD=2.1\%$). The critical interaction between *pronunciation* and *naming* was however not significant ($F(1,19)<1$, n.s., $\eta_p^2=.04$). Participants in the No Fixation condition looked more at the target in the post-naming phase regardless of pronunciation type, showing no evidence of mispronunciation detection.

The Fixation condition showed no significant main effect of *pronunciation* ($F(1,19)<1$, n.s., $\eta_p^2=.001$), but revealed an effect of *naming* ($F(1,19)=4.62$, $p=.045$, $\eta_p^2=.20$). The interaction between *pronunciation* and *naming* was not significant ($F(1,19)=1.48$, $p=.24$, $\eta_p^2=.07$), again showing no evidence of mispronunciation detection.

Finally the Fixation Plus condition showed a main effect of *pronunciation* ($F(1,19)=7.34$, $p=.014$, $\eta_p^2=.28$) with more looks at the target with a correct pronunciation (correct $M=53.9\%$, $SD=2.7\%$; incorrect $M=46.5\%$, $SD=3.0\%$). The effect of *naming* was significant ($F(1,19)=13.4$, $p=.002$, $\eta_p^2=.41$), and so was the interaction ($F(1,19)=10.17$, $p=.005$, $\eta_p^2=.5$). Simple effect analysis with Bonferroni correction confirmed an increase in looks at the target in the post-naming phase, only when words are correctly pronounced (correct $p=.002$, pre $M=45.9\%$, $SD=2.0\%$, post $M=61.9\%$, $SD=2.1\%$; incorrect $p=1.00$, pre $M=47.0\%$, $SD=2.3\%$ post $M=46.2\%$, $SD=3.5\%$). This condition showed the expected effect, namely better identification of the target image when the target was correctly produced as compared to an incorrect pronunciation.

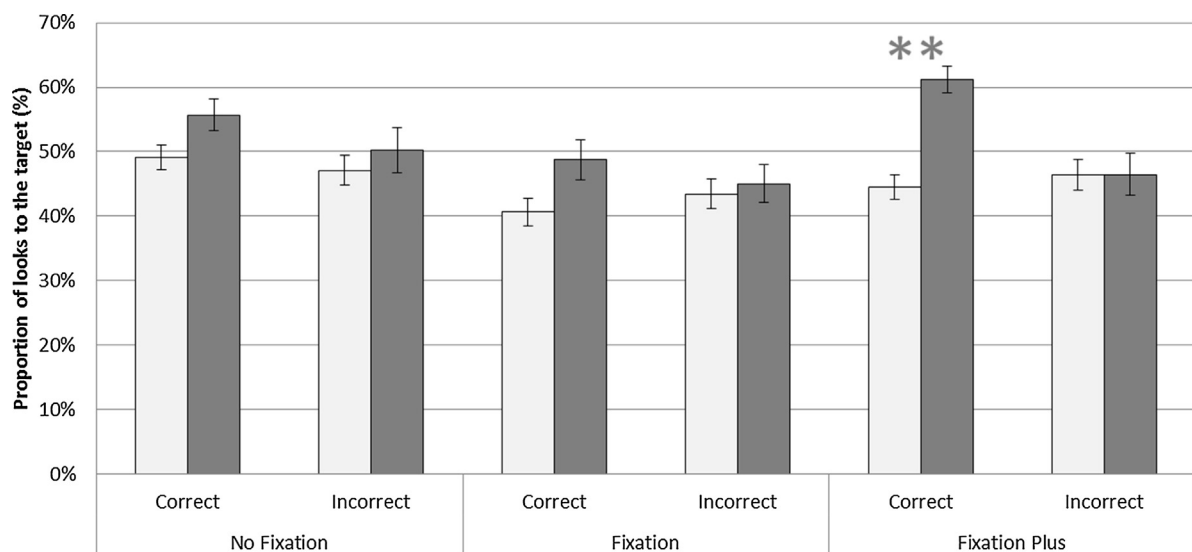


Fig. 2. Mean PTL for each condition and pronunciation type, for the 367–2000 ms time window, during pre-naming (light grey) and post-naming (dark grey). The double asterisk shows a difference with $p<.0001$. The lax criterion was used, and error bars represent 1 SE.

3.4.2. Comparison between dependent measures

So far, we ran a first series of analyses with *condition*, *pronunciation* and *naming* as factors, followed by a second series of analyses broken down by condition, with *pronunciation* and *naming* as factors. For the two dependent variables, we observed an overall agreement in the statistical tests, in particular for the *naming* variable, where the largest effect was expected overall. The more sensitive result, namely the interaction between *pronunciation* and *naming*, is clearly absent in the No Fixation and Fixation conditions, and robust in the Fixation Plus condition.

A rerun of all these analyses on trials retained with the strict attention criterion (that is, trials where children look at both pictures in the pre-naming phase) produced a comparable pattern of results (see Appendix D), except for the main effect of *pronunciation* that became weaker (from marginal to non-significant for the No Fixation condition, and from significant to marginal for the Fixation Plus condition).

In order to evaluate whether knowledge of the distracter word modulates the results, we re-ran the analysis on PTL measures for trials where the children knew both the target and the distracter, retaining most trials (No Fixation: 254/300; Fixation: 214/268; Fixation Plus: 246/299). We anticipated a loss of power, which was confirmed by the results in the first two conditions. There were no more significant effects in the No Fixation condition (*pronunciation*: $F(1,19) < 1$, n.s., $\eta_p^2 = .001$; *naming*: $F(1,19) = 1.98$, $p = .18$, $\eta_p^2 = .09$; interaction: $F(1,19) < 1$, n.s., $\eta_p^2 = .004$) and in the Fixation condition (*pronunciation*: $F(1,19) < 1$, n.s., $\eta_p^2 = .04$; *naming*: $F(1,19) = 1.54$, $p = .23$, $\eta_p^2 = .08$; interaction: $F(1,19) < 1$, n.s., $\eta_p^2 = .02$). In the Fixation Plus condition (note that one child did not contribute to the data), however, the previous results were replicated, with a main effect of *naming* ($F(1,18) = 13.2$, $p = .002$, $\eta_p^2 = .42$), that is, an increase between the pre-naming phase ($M = 45.4\%$, $SD = 1.8\%$) and the post-naming phase ($M = 53.0\%$, $SD = 1.9\%$). The effect of *pronunciation* was marginal (correct: $M = 50.9\%$, $SD = 3.1\%$; incorrect: $M = 47.2\%$, $SD = 2.7\%$; $F(1,18) = 1.59$, $p = .008$, $\eta_p^2 = .22$), but the interaction remained significant ($F(1,18) = 9.71$, $p = .006$, $\eta_p^2 = .35$).

It is not uncommon in the literature to analyze only the post-naming data (re. the Fernald and Swingle groups), which is the equivalent of running a post-hoc comparison in the ANOVA with *naming* and *pronunciation*. For the No fixation condition, there was a marginal effect of *pronunciation* in the post-naming phase ($F(1,19) = 3.30$, $p = .09$, $\eta_p^2 = .15$), with more looks to the target when the word is pronounced correctly. For the Fixation condition, this effect was not significant ($F(1,19) < 1$, n.s., $\eta_p^2 = .01$). Finally, for the Fixation Plus condition *pronunciation* lead to a significant difference ($F(1,19) = 12.97$, $p = .002$, $\eta_p^2 = .41$), as PTLs to the target increased when the word is pronounced correctly. Therefore in the current set of experiments, analyses focusing on the post-naming phase mirror those involving both pre- and post-naming phases, very likely because of infants' relatively balanced looking behaviour in the pre-naming phase (equal looks to the target and the distracter).

3.4.3. Time course analysis

Time course plots allowed us initially to visualize the time window where the naming effect was most likely taking place, ensuring that using the classic 367–2000 ms window would not miss out interesting data points. The following tests will provide us with a more precise measure as to when the two pronunciations elicit quantitatively different looking time behaviour. In order to analyze the time course of the proportions of look, we followed the methods advocated by von Holzen and Mani (2012), a non-parametric random permutation analysis (Maris & Oostenveld, 2007) to test effect of *pronunciation* across time, so as to identify the time period when looking times were significantly different. With this method however, and contrary to the classic PTL analysis including *naming* as a factor, pre-naming behaviour is not taken into consideration in relation to post-naming. This is why we considered another analysis of the time course, focusing on post-naming data corrected by the pre-naming data (PTL during the post-naming phase – average PTL in the whole pre-naming phase). In short, we ran two analyses, one similar to von Holzen and Mani (2012), and one where the average PTL to the target in the pre-naming phase (between 367 and 2000 ms) is subtracted from each data point. The latter will reveal any potential pronunciation effect that would have been masked by a pre-naming bias.

The procedure, best described in Maris and Oostenveld (2007), identifies the time period where looking behaviour differs between the correct and incorrect pronunciation, that is, the naming effect. In the first step of the procedure individual paired sample *t*-tests were performed at each time sample, and used to identify significant ($p < .05$) *t*-values. In step-two, clusters were identified by finding significant *t*-values that were contiguous across time. For each such cluster, a cluster-level *t*-value was calculated as the sum of all single sample *t*-values within the cluster. Analysis thereafter was based on these clusters and their associated cluster level *t*-value, rather than the individual (and highly non-independent) *t*-values. Since cluster level *t*-values could not be tested for significance against a standard *t* distribution, in step three of the procedure, the significance of each cluster was calculated by comparing its cluster-level *t*-value to a Monte Carlo distribution of cluster level *t*-values generated from the cluster with the largest cluster-level *t*-value. To do this each of the original paired sample *t*-tests that were used to generate this cluster were repeated, but with the data items of each pair randomly assigned between the two conditions. This was performed 1000 times to generate a Monte Carlo distribution of 1000 summed *t*-values corresponding to the null hypothesis. The summed *t*-values of these randomized tests provided a null distribution against which the actual cluster-level *t* statistic of each of the observed cluster could be compared. Thus, for each observed cluster, a Monte Carlo *p*-value was calculated as the proportion of the null distribution which had a cluster-level *t* statistic that exceeded the actual cluster-level *t*-statistic.

The first series of analyses were conducted on the whole trials (pre- and post-naming phases) and revealed no significant difference between correct and incorrect trials for the No Fixation condition (identified cluster: 900–1060 ms after target word onset; cluster *t* statistics = 12.13, Monte Carlo $p = .50$) and the Fixation condition (here, the two pronunciation conditions do not differ significantly from each other in any time window, therefore no Monte Carlo estimate was calculated) (Fig. 1a and b). For the Fixation Plus condition however (Fig. 1c), the two types of pronunciation differed significantly between 900 and 1580 ms post stimulus onset (cluster *t* statistics = 48.73, Monte Carlo $p = .04$), with an increase in looks at the target when it was correctly named.

The second series of analyses was conducted on the post-naming phase only, after subtracting the overall PTL to the target in the pre-naming phase. This is a similar approach to the inclusion of the salience score by Swingley and Aslin (2007), where they analyze PTL in the post-naming phase, minus PTL in the pre-naming phase; instead this time it is applied on each time frame. Again, no significant difference was obtained in the No fixation condition (all *t*-tests reveal $ps > .05$). In the Fixation condition a cluster between 2420 and 2460 ms post word onset was identified but was, however, non significant (cluster *t* statistics = 4.33, Monte Carlo $p = .96$). For the Fixation Plus condition, we found a significant difference window which was comparable to the first series of analyses, with more looks to the correctly pronounced target from 740 to 1940 ms after stimulus onset (cluster *t* statistic = 105.34, Monte Carlo $p = .002$).

The same analyses conducted on the dataset selected with the strict criterion reveal rather comparable results, with perhaps more sensitivity. On the whole duration of the trial, the No Fixation condition reveals difference cluster between 980 and 1700 ms after the word onset, but Monte Carlo simulations failed to confirm that this difference is significant (cluster *t* statistics = 12.76, $p = .53$). For the Fixation condition no cluster was identified (all *t*-tests reveal $ps > .05$). For the Fixation Plus condition, we replicate the significant effect of pronunciation, from 980 to 1700 ms after stimulus onset (cluster *t* statistics = 54.55, $p = .03$).

4. Discussion

The goal of the present study was to evaluate how infants' looking behaviour in the Inter-modal Preferential Looking paradigm varies as a function of noise generated by two simple methodological modifications, and how different methods of analysis best account for the resulting behavioural changes. Three groups of 20-month-olds were tested for recognition of correctly and incorrectly pronounced familiar words, in conditions that varied in terms of level of visual noise. In the first,

No Fixation condition, no central fixation stimulus was used and images were simply equated on size and judged as good exemplars, with no attempt to control for background and salience. In the second and third conditions (Fixation and Fixation Plus), a central fixation stimulus appeared between trials to attract the child's gaze to the centre before the onset of the next trial (and not at the onset of the post-naming phase as in Portillo et al., 2007, cited in Fernald et al., 2008). In addition, in the third, Fixation Plus condition, children were presented with images without any background and target-distracter pairs were matched for visual salience. To analyze the resulting data we examined the impact of different trial selection criteria, compared two dependent measures (LLK and PTL) and performed a time course analysis using a combination of Monte Carlo estimate and cluster analysis, to identify the time period where mispronunciation affected behaviour.

Following numerous studies using a similar paradigm (e.g., Mani & Plunkett, 2007; Swingley, 2003; Swingley & Aslin, 2000; White & Morgan, 2008), the expected result at 20 months was a naming effect restricted to the correctly pronounced targets, that is, an increase in looks to the target in the post-naming phase as compared to the pre-naming phase, only when the target word is pronounced correctly. We also expected that each methodological modification (addition of a central fixation and increased control of images) would contribute to enhance the quality of the data. The central point of the study was to determine which method of analysis would prove the most robust across methodological variations, and which would be the most sensitive.

Results overall revealed that all three groups showed a main effect of naming, that is, children fixated the target image longer after it was named. However, only the Fixation Plus group behaved as predicted by the literature, they showed a naming effect restricted to words correctly pronounced, just like in Mani and Plunkett (2007) or Swingley and Aslin (2000). Our main interpretation of these results is that the combination of the central fixation point and the selection of better-controlled images contributed to enhance the quality of data, and to reduce experimental noise, in the *pre-naming* phase, which in turn resulted in less variable *post-naming* data, as will be discussed below. We suspect that other parameters could act to reduce similarly the level of unwanted noise, such as the use of the same items to act as targets and distracters (as recommended by Fernald et al., 2008, see Mani & Plunkett, 2007; Swingley & Aslin, 2000) or the selection of the most frequent words in a child's vocabulary (e.g., Swingley & Aslin, 2000).

By progressively reducing the noise in the visual stimuli in the Fixation Plus condition, exploration of the visual stimuli during the pre-naming phase was more balanced with a PTL to the target around the expected 50%, despite a slight bias towards the distracter across all conditions. This bias must be due to reduced familiarity with the distracter objects (across all groups, participants were reported knowing the names of the target in 283 trials, and the name of the distracter in 246 trials), which thus worked as initial attracters. However, the pre-naming imbalance was not strictly comparable in the No Fixation (Fig. 1a) and the Fixation (Fig. 1b) conditions: whereas it was observed from the very onset of the pre-naming phase in the No Fixation condition, in the Fixation condition children looked equally long at targets and distracters from the onset of the pre-naming phase, and it is only after about 300 ms that the preference for distracters emerged. Given that the only methodological difference between the No Fixation and the Fixation conditions was the adjunction of a central fixation stimulus at the onset of each trial, it is quite likely that this central fixation point contributed to reducing the imbalance between target and distracter looks during the pre-naming phase. However, controlling for the quality of images had a cumulative positive effect, as seen in the Fixation Plus condition. Not only were pre-naming looks between targets and distracters more balanced, but the expected naming effect was obtained earlier, and was more robust than in the Fixation condition. What changed between the two conditions was a disappearance of the background and a quantitatively controlled balance in visual salience of the target-distracter pairs. Whether background control contributed more than saliency control to the sharpening of infants' behaviour remains undetermined in this study.

At this point, these results allow us to add to the recommendations of Fernald et al. (2008) and from the literature using the IPL paradigm: a fixation stimulus helps by centring the child's attention before trial onset, and carefully selected images even out the probability of looking at both images in the pre-naming phase, enhancing the sensitivity of the method.

Crucially, our central aim was to compare how these different methodological choices would impact on the robustness of data through the lens of different analytical choices such as the criteria for data selection and the dependent variables.

The literature shows that the criteria used for data inclusion or rejection in the pre-processing phase varies substantially across experimenters. A most reasonable practice – which we did not question – is to include only trials where the target word is known by the participant, as attested by parental questionnaire. More questionable is the practice of rejecting trials during which the child has not looked at both the distracter and the target at some point: does it have to be at some point during the entire trial (lax criterion), or during the pre-naming phase only (strict criterion)? We have shown that the two criteria, which measure the level of attentiveness during each trial, do not lead to fundamentally different results. Unsurprisingly more trials were rejected due to the application of the strict criterion (2.9% for the lax criterion vs. 11.0% for the strict criterion), resulting in a loss of experimental power. However, a close inspection of results in the PTL section in the Results section and Appendix D.2 shows that the size of the main effect of condition is larger when the strict criterion is applied. In contrast, applying the lax criterion results in larger sizes for all other effects, including pronunciation and naming as well as the crucial naming \times pronunciation interaction. This suggests that the overall behavioural adjustments due to methodological changes may be enhanced with the use of the strict criterion, but not the quality of the key effects (naming modulated by pronunciation). With the strict criterion, we ensure that children have seen the target and the distracter during the pre-naming phase. Upon hearing the label they would know that a mispronounced name does not correspond to any picture; they can then use a 'better match' strategy based on phonological overlap and look slightly longer at the target. This translates into a relative decrease in the size of the pronunciation effect as compared to the same data analyzed with the

lax criterion. In contrast, with the lax criterion, we also include trials where the child has only checked the target during pre-naming,¹ and therefore, can reasonably assume that the mispronunciation can refer to the unchecked item. This results in a slightly higher number of looks towards the distracter during the post-naming phase. To sum up, data may suggest that we do not measure the same behaviour or strategy if we apply the strict or the lax criterion: for the former, we may measure a better-fit strategy based on the degree of phonological overlap, whereas in the latter, children may produce a response based on a Mutual Exclusivity-type principle (Halberda, 2003). If this speculative assumption was corroborated by further research, this should be kept in mind when deciding for one criterion over the other, depending on the theoretical goals of the experiment.

We have seen that it is common practice to exclude trials in which children do not know the name for the target object; is it justified to also exclude trials where the child does not know the distracter (as done by, among others, Swingley et al., 1999)? On the one hand, this could have some advantage: the children would be more likely to look away from an incorrectly named target image and attach the mispronounced label to the distracter (White & Morgan, 2008), strengthening the mispronunciation detection effect. On the other hand, unknown distracters could result in children showing a familiarity effect rather than a naming effect (looking at the named target simply because they have a name for it, not because it has been named with its specific label). Quite pragmatically, excluding trials in which the child does not know the distracter would possibly result in a loss of experimental power. This was indeed the case for the No Fixation and Fixation conditions, but not in the more robust condition where the critical interaction was replicated. Therefore it appears that the application of this criterion does not substantially modify the quality of the data, at least not in the current study. It should be kept in mind however that, similarly to what was discussed above for the use of the lax vs. strict criteria, knowing, or not knowing, the distracter label may modify the strategy that the child uses in the procedure. An unknown distracter promotes the use of the Mutual Exclusivity principle whereas a known distracter encourages the use of a best-match strategy based on the degree of phonological overlap.

Regarding the choice of the dependent variable, the literature often presents side by side analyses based on the proportion of looks to the target (PTL) and on the longest look to the target (LLK), as they usually show similar results. The same conclusion can be applied here, although with a caveat. A close inspection of statistics in the result section shows that in most analyses, effect sizes are larger for the LLK measures than for the PTL ones. This could be due to the fact that in the vast majority of cases, the longest look is also the first look towards the target, and during that period which lasts about 700 ms (see Table 1), the child computes all the information that is needed to correctly identify the target. Possibly all further looks towards the target are either verification or random noise, which is incorporated in the PTL measure but not in the LLK variable, resulting in less variable data in the latter than in the former measure.

Finally, we questioned the importance of adjusting the time window of analysis, and investigated the relevance of a time course analysis. Depending on the age and/or vocabulary size of the participants, it is common practice to adjust the onset of the post-naming phase to account for variation in gaze shift latency (Fernald et al., 2008). In addition, many factors can influence the processing time of the target and distracter pictures, starting with the nature and complexity of the auditory stimulus, the visual properties of the stimuli (e.g., Arias-Trejo & Plunkett, 2010), or the type of distracter (familiar vs. unfamiliar; White & Morgan, 2008). Therefore a time window fixed a priori may not be the most accurate. Of course, selecting for each experiment a time window based on the visual inspection of the data would be unacceptable as it would lead to a strong human bias. One way around this is to generalize the use of the time course analysis as reported here which allows the identification of time windows where the naming or pronunciation effects are indeed significant. It seems that the statistical analysis of the time course provided an accurate estimate of looking behaviour, since it allowed us to distinguish between a very short lived pronunciation effect (in the No Fixation condition) and an enduring one (with the Fixation Plus condition). This mirrored the outcome of the classic mean-based looking times analyses, namely a robust interaction pronunciation \times naming in the Fixation Plus condition and none in the No Fixation condition. While very promising to estimate the speed of word recognition (like Durrant et al., 2014; Fernald et al., 2006), this approach needs to evolve to establish the minimal temporal window where pronunciation differences are meaningful (re. the very short lived pronunciation effect in the No Fixation condition).

In summary, observations based on infants' behaviour in a classic IPL task can vary quite substantially depending on the methodological parameters chosen during the pre-processing period or data analysis. Perhaps the vulnerability of the data are best illustrated in Fig. 2c which displays the results of the Fixation Plus condition. Correctly and incorrectly pronounced words produced different looking times for about 700 ms, as revealed by the time course analysis. This is a rather short window of interest as compared to the entire duration of the trials, for example as compared to head-turn procedures which typically generate differences of about 2 s of looking times between conditions (e.g., Mattys, Jusczyk, Luce, & Morgan, 1999; Jusczyk & Aslin, 1995). It can be argued that IPL is a more direct and precise measure of auditory processing than head turn paradigms as it does not rely on an experimenter's intervention during the session (whereas head turn set-ups usually do: Floccia et al., 2012; Nazzi, Jusczyk, & Johnson, 2000; Schmale & Seidl, 2009). Yet this augmented precision perhaps makes the IPL tool more prone to vary with methodological noise. To borrow an example from physical instruments, a digital thermometer might be more precise than a mercury one, yet thanks to its inertia the latter is more likely to give robust

¹ The number of trials in which the children only look at the target during the pre-naming phase is as follows: No Fixation: 21/285 trials; Fixation: 19/270; Fixation Plus: 15/293.

repeated measures than the former. It is our hope that this methodological study will contribute to sharpen the use of this invaluable paradigm in the quest of infants' representation and processing of visual and auditory information.

Acknowledgment

Part of the study was funded by an ESRC grant 2009/2011 RES-000-22-3596 and one from the Nuffield Foundation, both awarded to the fourth author. The authors thank William A. Simpson for his advice on saliency measures. Samantha Durrant is a Post-Doctoral Research Associate in the International Centre for Language and Communicative Development (LuCiD) at the University of Liverpool. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

Appendix A.

Characteristics of some IPL studies, ordered by type of task then by alphabetical order of the authors. *T* and *D* stand for looks at the target and the distracter respectively. Unless specified differently, the target word is presented in a carrier sentence and uttered only once.

Meints et al. (2002)	Task	Comprehension of prepositions
	Age(s)	15, 18, 24 months
	Fixation point	Light in centre before trial onset
	Trial structure	5.5 s, onset sentence at 600 ms, onset target word at 2750 ms (pre-naming 0–2750 ms, post-naming 2750–5500 ms)
	Exclusion	Presented only with words they know (2004)
	Analyses	Longest look (<i>T</i> – <i>D</i> , pre- and post-naming) Total looking time
Dittmar et al. (2008)	Task	Verb comprehension and learning
	Age(s)	21 months
	Fixation point	Not before test phase
	Trial structure	Test phase lasts 8 s
	Exclusion	Inattentive trials
	Analyses	Proportions of looks per 2 s and overall
Gurteen et al. (2011)	Task	Word learning (with real objects)
	Age(s)	13, 17 months
	Fixation point	Centre light before naming onset
	Trial structure	Only from target word offset
	Exclusion	Familiar words are reported as known
	Analyses	Proportion of looks ($T/(T + D + \text{other looks}) * 100$) Comparison with chance at 50%
Schafer and Plunkett (1998)	Task	Word learning
	Age(s)	12–17 months
	Fixation point	Light in centre before trial onset
	Trial structure	Test phase: pre-naming 0–3 s, post-naming 3–7 s, carrier sentences with 3 tokens for <i>T</i>
	Exclusion	Ignore pre-naming
	Analyses	Total looking time Longest look Graph for time course of proportions (no statistics) Shift latency
Swingley and Aslin (2007)	Task	Word learning
	Age(s)	19 months
	Fixation point	None
	Trial structure	Test phase: 6.5 s, pre-naming 0–3 s, post-naming 3–6.5 s, with two sentences containing <i>T</i>
	Exclusion	Only data 367–2000 ms post-naming
	Analyses	Proportion of looks ($T/(T + D)$) Saliency score (proportions of looks to <i>T</i> in pre-naming) Difference proportion in post-naming minus saliency score
Mani and Plunkett (2008)	Task	Word learning (with mispronunciations)
	Age(s)	14, 18 months
	Fixation point	None
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming 2500–5000 ms, word onset at 2500 and 4000 ms
	Exclusion	Only data 367–2500 ms post-naming Only trials where both pictures are fixated

Arias-Trejo and Plunkett (2009)	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T+D)$) Latency of switch
	Task	Priming (semantic)
	Age(s)	18, 21 months
	Fixation point	Stimulus in the centre before onset of prime word
	Trial structure	Carrier sentence finishing with prime, T word onset 200 ms after offset prime, images appear 200 ms after offset T word, for 2500 ms
	Exclusion	Look away trials are excluded Unfamiliar words are excluded
Mani and Plunkett (2011a, 2011b)	Analyses	On all data 0–2500 ms Longest look ($T - D$) Proportion of looks ($T/(T+D)$)
	Task	Priming (phonological)
	Age(s)	24 months
	Fixation point	None, but prime image is centred
	Trial structure	Prime image for 1.5 s, blank 200 ms, two images for 2.5 s, naming 50 ms after onset of images
	Exclusion	Only data 233–2000 ms post-naming Only trials where child knows T and prime
Houston-Price et al. (2010)	Analyses	Proportions of looks ($T/(T+D)$)
	Task	Mutual exclusivity
	Age(s)	17–22 months
	Fixation point	None
	Trial structure	7 s, pre-naming 0–3.5 s, post-naming at word onset 3.5–7 s
	Exclusion	Only children who knew the 3 familiar words
Fernald et al. (1998)	Analyses	Proportion of looks ($T/(T+D)$)
	Task	Word recognition
	Age(s)	15, 18 and 24 months
	Fixation point	None
	Trial structure	6 s, pre-naming 0–4 s, post-naming 4–6 s from sentences onset (target word repeated once)
	Exclusion	Final analysis excludes look away trials during word naming
Fernald et al. (2006)	Analyses	Proportions of looks to T and D during 4 s after word offset Latency of shift to T Graph for time course of proportions (no statistics)
	Task	Word recognition
	Age(s)	15, 18, 21 and 24 months
	Fixation point	None
	Trial structure	Pre-naming 0–3 s, post-naming 3–6 s or 3–8 s
	Exclusion	Only data 300–1800 ms post-naming Excluding outlier gaze shifts Excluding participants not contributing to all conditions
Houston-Price et al. (2007)	Analyses	Latency of shift to T Proportion of shifts to T (300–1800 ms post-naming) Multiple window analysis Graph for time course of proportions (no statistics)
	Task	Word recognition (familiarity vs. novelty)
	Age(s)	15, 18, 21 months
	Fixation point	Light in centre before trial onset
	Trial structure	5 s, pre-naming 0–2600 ms, post-naming at T word onset 2600–5000 ms
	Exclusion	None
Meints et al. (1999)	Analyses	Proportion of looks ($T/(T+D)$)
	Task	Word comprehension
	Age(s)	12, 18, 24 months
	Fixation point	Light in centre before trial onset
	Trial structure	5.5 s, onset sentence at 600 ms, onset target word at 2100 ms (pre-naming 0–2100 ms, post-naming 2100–5500 ms)
	Exclusion	Presented only with words they know Analyses Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T+D)$)
Meints et al. (2004)	Analyses	Proportion of looks ($T/(T+D)$)
	Task	Word comprehension
	Age(s)	14, 18
	Fixation point	Light in centre before trial onset

Styles and Plunkett (2008)	Trial structure	5.5 s, onset sentence at 600 ms, onset target word at 2750 ms (pre-naming 0–2750 ms, post-naming 2750–5500 ms)
	Exclusion	Presented only with words they know
	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$)
	Task	Word recognition
	Age(s)	18 months
Swingley et al. (1999)	Fixation point	Between trials, distracter sentence played in the centre
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming 2500–5000 ms
	Exclusion	2nd analysis with only participants contributing with data in all conditions
	Analyses	Longest look difference ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$) Two sets of data, broad and narrow criteria
	Task	Word recognition
Arias-Trejo and Plunkett (2010)	Age(s)	24 months
	Fixation point	None
	Trial structure	9 s, pre-naming 4 s, post-naming 5 s
	Exclusion	All words are known
	Analyses	Proportion of looks ($T/(T + D)$) Latency shift to T , 233–2000 ms post-naming Graph for time course of proportions (no statistics)
Bailey and Plunkett (2002)	Task	Word recognition (mispronunciations)
	Age(s)	18, 21, 24 months
	Fixation point	None
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at T onset, 2500–5000 ms
	Exclusion	Only trials where T is familiar Inattentive trials are excluded
Ballem and Plunkett (2005)	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$) Graph for time course of proportions (no statistics) Latency (360–1800 ms)
	Task	Word recognition (mispronunciations)
	Age(s)	18, 24 months
	Fixation point	None
	Trial structure	5.5 s, pre-naming 0–2750 ms, word onset at 2750 ms, post-naming 3150–5500 ms
Durrant et al. (2014)	Exclusion	Only trials where minimum of 1500 ms of looks in each phase
	Analyses	Longest look difference ($T - D$)
	Task	Word recognition (mispronunciation)
	Age(s)	14 months
	Fixation point	None
Mani et al. (2008)	Trial structure	Pre-naming 0–2600 ms, post-naming from word onset 2600–5000 ms
	Exclusion	Only trials with looks at least 1500 ms long in each phase Only data 200–2400 post-naming
	Analyses	Proportions of looks ($T/(T + D)$)
	Task	Word recognition (mispronunciations)
	Age(s)	20 months
Mani and Plunkett (2007)	Fixation point	Between trials
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at T onset 2500–5000 ms
	Exclusion	Only data 367–2000 ms post-naming Only trials where T is familiar Only trials where both pictures are fixated
	Analyses	Proportion of looks ($T/(T + D)$) Graph for time course of proportions (no statistics)
	Task	Word recognition (mispronunciations)
Mani and Plunkett (2007)	Age(s)	18 months
	Fixation point	None
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at T onset 2500–5000 ms
	Exclusion	Only data 367–2500 ms post-naming Only trials where both pictures are fixated
	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$)
Mani and Plunkett (2007)	Task	Word recognition (mispronunciations)
	Age(s)	15, 18, 24 months
Mani and Plunkett (2007)	Fixation point	None
	Analyses	Longest look difference ($T - D$)

	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at <i>T</i> onset 2500–5000 ms
	Exclusion	Only data 367–2500 ms post-naming 15 months: only trials where look at <i>T</i> and <i>D</i> during whole trial 18, 24 months: only trials where look at <i>T</i> and <i>D</i> during pre-naming
	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$) Percentage of switch to <i>T</i> when fixating <i>D</i> at word onset (no statistics) Latency of switch
Mani and Plunkett (2010)	Task	Word recognition (mispronunciations)
	Age(s)	12 months
	Fixation point	None
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at <i>T</i> onset 2500–5000 ms
	Exclusion	Only trials where look at <i>T</i> and <i>D</i> during pre-naming
	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$)
Mani and Plunkett (2011a, 2011b)	Task	Word learning (mispronunciations)
	Age(s)	18, 24 months
	Fixation point	None
	Trial structure	5 s, pre-naming 0–2500 ms, post-naming at <i>T</i> onset 2500–5000 ms
	Exclusion	Only trials where look at <i>T</i> and <i>D</i> during pre-naming
	Analyses	Longest look ($T - D$, pre- and post-naming) Proportion of looks ($T/(T + D)$)
Ramon-Casas et al. (2009)	Task	Word recognition (mispronunciations)
	Age(s)	17–24 months
	Fixation point	None
	Trial structure	Pre-naming 0–2 s, sentence, post-naming 0–2 s after sentence offset
	Exclusion	Only data 360–2000 ms post-naming
	Analyses	Proportion of looks ($T - D$) Separate analyses re. image fixed at onset Plots for individual averages
Swingley (2003)	Task	Word recognition (mispronunciations)
	Age(s)	18–20 months
	Fixation point	None, but first trial initiated when fixating animated film displayed on the screen
	Trial structure	6.5 s, pre-naming 0–2.5 s, post-naming 2.5–6.5 s
	Exclusion	Only data 360–2000 ms post-naming
	Analyses	Proportion of looks ($T/(T + D)$) Difference in proportion of looks $T - D$
Swingley (2007)	Task	Word learning (with mispronunciations)
	Age(s)	18–20 months
	Fixation point	None
	Trial structure	Pre-naming 0–2 s, post-naming at word onset 2–5.5 s
	Exclusion	Only data 360–2000 ms post-naming
	Analyses	Proportion of looks (vs. chance, in post-naming)
Swingley and Aslin (2000)	Task	Word recognition (mispronunciations)
	Age(s)	18–23 months
	Fixation point	None
	Trial structure	9 s, pre-naming 0–3 s, post-naming at sentence onset 3 s, sentence repeated once
	Exclusion	Only data 367–2000 ms post-naming
	Analyses	Proportion of looks ($T/(T + D)$) Latency of shift to <i>T</i> Second analysis with productive CDI Graph for time course of proportions (no statistics)
Swingley and Aslin (2002)	Task	Word recognition (mispronunciations)
	Age(s)	14, 15 months
	Fixation point	None, but first trial initiated when fixating animated film displayed on the screen
	Trial structure	Pre-naming 0–3 s, post-naming at onset of first of two sentences
	Exclusion	Only data 367–2000 ms post-naming
	Analyses	Proportion of looks ($T/(T + D)$)
White and Morgan (2008)	Task	Word recognition (mispronunciation)
	Age(s)	19 months
	Fixation point	Light in centre before trial onset and post-naming phase

Trial structure	13 s, pre-naming 4 s, post-naming 9 s (2 carrier sentences) for Exp. 1, post-naming 5 s for Exp. 2 (1 carrier sentence)
Exclusion	Outlier participants (3)
Analyses	Data only for the first 3 s of post-naming Proportion of looks ($T/(T+D)$) Difference Proportion of looks minus salience in pre-naming

































Appendix B.

List of the target words along with their phonetic transcription, the transcription of vowel and consonant mispronunciations and their corresponding distracter object.

Target	Vowel misp.	Consonant misp.	Distracter	
Ball	/bɔ:l/	/bu:l/	/gɔ:l/	bear
Bath	/bɑ:θ/	/bεθ/	/dɑ:θ/	boat
Bed	/bɛd/	/b^d/	/pɛd/	book
Bib	/bɪb/	/bæb/	/dɪb/	boot
Bread	/brɛd/	/brɔ:d/	/grɛd/	brush
Bus	/b^s/	/bæs/	/p^s/	bike
Cat	/kæt/	/kɑ:t/	/gæt/	cow
Cot	/kɒt/	/kɔ:t/	/tɒt/	car
Cup	/k^p/	/kɛp/	/g^p/	clock
Dog	/dɒg/	/dɒg/	/bɒg/	duck
Foot	/fɒt/	/fɔ:t/	/sɒt/	fish
Hat	/hæt/	/hɛt/	/fæt/	horse
Keys	/ki:z/	/kæz/	/ti:z/	coat
Shoe	/ʃu:/	/ʃi:/	/fu:/	shop
Sock	/sɒk/	/sɒk/	/zɒk/	spoon
Tree	/tri:/	/tru:/	/pri:/	train

Appendix C.

List of the visual stimuli for the No Fixation (the list is identical to that for the Fixation condition) and the Fixation Plus condition.

Target	No Fixation	Fixation Plus	Distracter	No Fixation	Fixation Plus
Ball			bear		
Bath			boat		
Bed			book		
Bib			boot		
Bread			brush		
Bus			bike		
Cat			cow		
Cot			car		

Cup



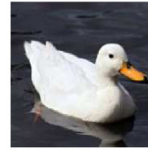
clock



Dog



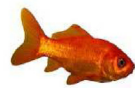
duck



Foot



fish



Hat



horse



Keys



coat



Shoe



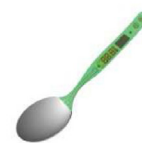
shop



Sock



spoon



Tree



train



Smiley face (fixation point)



Appendix D.

Replication of the analyses (descriptive statistics and ANOVAs) with the strict criterion of trial inclusion (whereby only trials with looks at both the target and the distracter in the pre-naming phase are included).

D.1. Mean PTL to the target (in %) and LLK to the target and the distracter (in ms) in the pre- and post-naming phases, for all conditions (standard error in brackets).

Condition	PTL to target			LLK to target			LLK to distracter		
	Pre-naming	Post-naming		Pre-naming	Post-naming		Pre-naming	Post-naming	
		Correct	Incorrect		Correct	Incorrect		Correct	Incorrect
No fixation	48.1 (1.9)	54.4 (2.5)	54.9 (3.8)	685.4 (34.4)	783.9 (48.8)	620.4 (42.2)	735.8 (30.8)	803.3 (73.4)	628.6 (52.9)
Fixation	41.1 (1.0)	48.2 (3.2)	46.1 (3.2)	581.4 (15.6)	693.8 (56.7)	724.1 (45.2)	856.1 (20.0)	585.0 (19.6)	763.0 (55.1)
Fixation plus	47.8 (1.2)	62.6 (2.5)	47.9 (4.5)	676.7 (19.7)	927.9 (52.9)	497.7 (28.1)	741.1 (20.9)	648.0 (61.7)	726.9 (78.3)

D.2. Statistics (F , p and partial η_p^2 values) for the factors condition, pronunciation, naming and their interactions. Dependent variables are the LLK and the PTL to the target. Significant effects are indicated in bold, marginal effects ($p < .10$) in italics. p values where significance levels have changed compared to the lax criterion are underlined

Factor		PTL	LLK
Condition	F	8.65	6.14
	p	.001	.004
	η^2	.23	.18
Pronunciation	F	2.39	5.19
	p	<u>.13</u>	.03
	η^2	.04	.08
Naming	F	20.96	14.87
	p	<.0001	<.0001
	η^2	.27	.21
Condition * Pronunciation	F	1.48	2.45
	p	.24	.10
	η^2	.05	.08
Condition * Naming	F	<1	<1
	p	n.s.	n.s.
	η^2	<.01	<.01
Pronunciation * Naming	F	2.79	3.33
	p	<u>.10</u>	<u>.07</u>
	η^2	.05	.06
Condition * Pronunciation * Naming	F	3.15	4.3
	p	.05	.02
	η^2	.10	.13

D.3. Statistics (F , p and partial η_p^2 values) for the factors pronunciation, naming and their interaction, broken down by condition (No fixation, Fixation and Fixation Plus). Dependent variables are the Longest Look and the Proportion of Looks to the Target. Significant effects are indicated in bold, marginal effects ($p < .10$) in italics. p values where significance levels have changed compared to the lax criterion are underlined.

	PTL			LLK		
	No fixation	Fixation	Fixation Plus	No fixation	Fixation	Fixation Plus
<i>Pronunciation</i>						
F	<1	<1	3.86	<1	<1	7.69
p	n.s.	n.s.	<u>.06</u>	n.s.	n.s.	.01
η_p^2	<u>.03</u>	0.0006	0.17	0.02	<.01	0.29
<i>Naming</i>						
F	5.51	7.59	10.05	4.44	4.62	6.14
p	.04	.01	.005	.05	.05	.002
η_p^2	0.21	0.29	0.35	0.19	0.2	0.24
<i>Interaction</i>						
F	<1	1.68	7.3	<1	2.02	8.91
p	n.s.	0.21	.01	n.s.	0.17	.008
η_p^2	0.02	0.08	0.28	0.03	0.1	0.32

References

- Althaus, N., & Mareschal, D. (2012). Using saliency maps to separate competing processes in infant visual cognition. *Child Development*, 83., (4), 1122–1128.
- Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society*, 364., 3633–3647.
- Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, 105., 63–80.
- Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effect in the infant lexicon. *Cognition*, 128., 214–227.
- Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17., 1265–1282.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32., 159–173.
- Bastiaansen, M. C. M., van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17., (3), 530–541.
- Chinga, G., & Syverud, K. (2007). Quantification of paper mass distributions within local picking areas. *Nordic Pulp and Paper Research*, 22., 441–446.
- Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, 105., 376–385.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). Young German children's early syntactic competence: A preferential looking study. *Developmental Science*, 11., 575–582.
- Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2014). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, 21., 1–19. <http://dx.doi.org/10.1017/S0305000914000063>
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42., 1. <http://dx.doi.org/10.1037/0012-1649.42.1.98>
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9., 228–231.
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). New York: John Benjamins.
- Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, 124., 95–100.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14., 23–45.
- Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, 109., 201–217.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38., 1531–1549.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87., B23–B34. [http://dx.doi.org/10.1016/S0010-0277\(02\)00186-5](http://dx.doi.org/10.1016/S0010-0277(02)00186-5)
- Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, 27., 689–705.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58., (1), 78–80.
- Houston-Price, C., Caloghris, Z., & Raviglione, E. (2010). Language experience shapes the development of mutual exclusivity bias. *Infancy*, 15., 125–150.
- Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, 34., 701–724.
- Jusczyk, P., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 1–23.
- Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, 61., 1–5. <http://dx.doi.org/10.3389/fnhum.2011.00061>
- Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological specificity of vowel contrasts at 18-months. *Language and Speech*, 51., 3–21.
- Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, 66., (4), 612–622. <http://dx.doi.org/10.1016/j.jml.2012.03.003>
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57., 252–272.
- Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, 11., 53–59.
- Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps or tups. *Infancy*, 15., 445–470.
- Mani, N., & Plunkett, K. (2011a). Does size matter? Sub-segmental cues to vowel mispronunciation detection. *Journal of Child Language*, 38., 606–627.
- Mani, N., & Plunkett, K. (2011b). Phonological priming and cohort effects in toddlers. *Cognition*, 121., 196–206.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164., 177–190.
- Mather, E., & Plunkett, K. (2011). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language*, 38., 933–950.
- Mattys, S. L., Jusczyk, P., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38., 465–494.
- Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology*, 35., 1072–1078.
- Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2002). What is 'on' and 'under' for 15-, 18- and 24-month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology*, 20., 113–130.
- Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2004). The cow on the high street: Effects of background context on early naming. *Cognitive Development*, 19., 275–290.
- Meints, K., & Woodford, A. (2008). *Lincoln Infant Lab Package 1.0: A new programme package for IPL, Preferential Listening, Habituation and Eyetracking*. Retrieved from <http://www.lincoln.ac.uk/psychology/babylab.htm>
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43., 1–19. <http://dx.doi.org/10.1006/jmla.2000.2698>
- Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of Acoustical Society of America*, 100., 577–592.
- Ramon-Casas, M., Swingle, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, 59., 96–121.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Verderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33., 217–236.
- Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69., 309–320.
- Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and accent: Flexibility of early word representation. *Developmental Science*, 12., (4), 583–601. <http://dx.doi.org/10.1111/j.1467-7687.2009.00809.x>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86., (2), 420–428.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human, Perception and Performance*, 6., 174–215.
- Stager, C., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than word learning tasks. *Nature*, 388., 381–382.

- Styles, S. J., & Plunkett, K. (2008). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*: 36., (4), 895–908. <http://dx.doi.org/10.1017/S0305000908009264>
- Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition*: 1., (1), 1–24. <http://dx.doi.org/10.1515/LANGCOG.2009.001>
- Swingle, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*: 46., 265–294.
- Swingle, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*: 43., 454–464.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*: 76., 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*: 13., 480–484.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*: 54., 99–132.
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*: 71., 73–108.
- von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*: 113., 569–586.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*: 3., 1–30.
- Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting*: 6., 171–178.
- Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*: 37., 278–286.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*: 7., 49–63.
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*: 59., 114–132.
- White, K. S., Morgan, J. L., & Wier, L. M. (2005). When is dar a car? Effects of mispronunciation and referential context on sound-meaning mappings. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th annual Boston University conference on language development*. Somerville, MA: Cascadia Press.
- Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development*: 3., 199–231.
- Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*: 6., 179–208.